SECURITY CLASSIFICATION OF THIS PAGE

(2)

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1 AD-A205 585	1b. RESTRICTIVE MARKINGS
2. NA	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NA	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 89-0276
6a. NAME OF PERFORMING ORGANIZATION: Penn State University	6b. OFFICE SYMBOL (If applicable):
6c. ADDRESS (City, State, and ZIP Code): Department of Statistics, 123 Pond Laboratory, University Park, PA 16802	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM
	7b. ADDRESS (City, State, and ZIP Code): Bldg. 410, Bolling AFB, DC 30332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION: AFOSR	8b. OFFICE SYMBOL (If applicable): NM
	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-88-0030
8c. ADDRESS (City, State, and ZIP Code): AFOSR/NM, Bldg. 410, Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 61102F; PROJECT NO. 2304; TASK NO A5; WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

MULTIVARIATE ANALYSIS AND ITS APPLICATIONS

12. PERSONAL AUTHOR(S)

C. R. Rao

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Report	FROM 1 Oct 87 TO 31 Dec 88	14 February 1989	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

During the period of October 1, 1987 - December 31, 1988, research was carried out in several new areas of multivariate analysis of interest to the Air Force. They have applications in manufacturing technology, automation, expert systems, pattern recognition and machine intelligence.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. Eytan Barouch	22b TELEPHONE (Include Area Code): (202) 767-4940	22c. OFFICE SYMBOL: NM

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

MULTIVARIATE ANALYSIS AND ITS APPLICATIONS

Annual Report covering the period
October 1, 1987 - December 31, 1988
under the contract AFOSR-88-~~0300~~ 0030

CENTER FOR MULTIVARIATE ANALYSIS
Department of Statistics
123 Pond Laboratory
Penn State University
University Park, PA 16802

MULTIVARIATE ANALYSIS AND ITS APPLICATIONS

During the period of October 1, 1987 - December 31, 1988, research was carried out in several new areas of multivariate analysis of interest to the Air Force. They have applications in manufacturing technology, automation, expert systems, pattern recognition and machine intelligence.

About 59 Technical Reports were issued for publication in journals and presenting at conferences. A list of the Technical Reports together with the abstracts is given in the Appendix to this report. A brief outline of some of the important contributions is given below.

1. L_1-NORM IN MULTIVARIATE STATISTICAL ANALYSIS

The classical methods of multivariate analysis are based on the averages and variances and covariances computed from the sample data; the underlying theory is based on the least squares technique using the L_2-norm. The estimates so obtained are not robust in the presence of outliers, recording errors and deviations from normality. A new methodology based on the L_1-norm, which is more robust, is developed.

The joint asymptotic distribution of the marginal medians is obtained as a basis for inference on the unknown median values (or means for symmetrical populations). All the classical tests based on the averages have been reformulated in terms of the medians. The nuisance parameters in the distribution are efficiently estimated using a new method of quantile density estimation, and used to adjust the test procedures.

Asymptotic inference procedures on the regression parameters based on the L_1-norm are developed in the univariate case and methods for eliminating nuisance parameters are discussed. The results are extended to the multivariate case.

Haldane defined what is called a spatial median of a set of observed vectors by minimizing the sum of the distances of the observed vectors from a fixed vector. The optimum fixed vector so computed is called the spatial median. This concept is extended to the estimation of regression parameters in a multivariate linear model. The sampling theory of such estimates and the tests based on them are developed.

The efficiencies of the estimates computed from the L_1-norm are compared with those of the L_2-norm (least squares). The robustness of the inference procedures based on the L_1-norm is examined.

A review is made of the previous work on M-estimation and some of the deficiencies in the proofs and assumptions have been corrected. This has led to the development of a unified theory of M-estimation in a rigorous way. Further work in this area is in progress.

2. MODEL SELECTION

The work on model selection is continued during the period under review. For purposes of predicting future values it is important to know the underlying model (probability mechanism). The exact model in a given situation, such as in a regression problem, time series, growth studies, logistic regression or a control system, is usually unknown. Then the question arises as to how a model can be selected on the basis of observed data? A very general criterion was developed at the Center for Multivariate Analysis for this purpose, which involves the maximization of the log likelihood of the observations after subtracting a penalty, which is a function of the number of unknown parameters in the model and the

sample size. Although the form of the penalty function was established, the exact inputs for a particular choice in a given situation remained to be investigated.

A number of studies have been carried out with special reference to

* choice of variables in a regression problem,
* dimensionality reduction in multinomial logistic regression model,
* order of an autoregressive time series,
* order of an ARIMA process.

Some guidelines have been provided on the basis of theoretical studies and extensive simulations.

3. CHARACTERIZATION OF PROBABILITY DISTRIBUTIONS

Characterization of probability distributions is important in data analysis as well as in studying the underlying structure of a random variable. Several important contributions have been made in this area.

Characterizations have been obtained for a univariate normal distribution through independence of linear statistics and constancy of the regression of a polynomial of sample average on residuals.

The structure of elliptically symmetric distributions have been investigated through the notion of exchangeability.

Further work has been done on the problem of the integrated Cauchy functional equation which plays an important role in a variety of problems, such as reliability theory, study of order statistics and sequential analysis.

Characterization theory is basic to problems of statistical inference in that it enables us

* to detect departures from a specified distribution,
* to choose appropriate estimates for parameters,
* to select efficient test procedures.

4. DISCRIMINANT ANALYSIS

The problem of identifying an individual as a member of a particular class among a set of possible classes, on the basis of observations taken on the individual, is of great importance in research as well as in routine operations. For instance, one may ask whether an object (say, a plane flying in the sky or a submarine under water) belongs to a given category (friendly or enemy). We can take a given set observations on the object and take a decision. This is not necessarily an efficient way, specially if the loss due to wrong decisions has to be controlled at a given low level. A new method is developed in which observations are made sequentially and a decision is taken when sufficient evidence is available. The advantage of this method is that the cost of making observations and analysing data can be made a minimum while controlling the loss due to wrong decisions.

In another investigation the linear discriminant function is shown to be admissible in a larger class of spherical distributions.

Tests for redundancy of variables in discriminant analysis have been studied by a number of authors. These tests have been extended to include redundancy in covariates besides the main variables.

5. SELECTION OF THE BEST POPULATION

Suppose that there is a set of populations with unknown mean values and some nuisance parameters, and we have a sample of observations from

each population. The problem is to select the best population, i.e., with the largest mean, or select a subset of populations which contains the best population. Since decisions are made on the basis of sample data, they will be subject to error. Considerable research was done in this area during the last 30 years.

A new method is introduced which is sequential in nature. Observations are made sequentially and decision is taken at each stage to terminate sampling and make a selection or continue sampling. An optimum sequential rule is provided to guarantee that with a given probability the best population is included in the selected subset and each selected population is within some fixed distance from the best population.

6. LINEAR MODELS WITH MIXED EFFECTS

Linear models with fixed effects have been studied extensively over the last fifty years, but not much work is done on mixed effects models, i.e., with random and fixed effects. A unified approach is developed for the estimation of fixed effects, random effects and random error in a mixed effects Gauss-Markoff model. The expressions for the estimators and the mean square errors are obtained in a general situation without making any assumption on the ranks of the matrices involved. A new concpet of conditioned equations (similar to normal equations) is introduced for the simultaneous estimation of mixed effects and random error. The methods developed for mixed effects models are similar to those for fixed effects models, thus providing a unified theory.

The geometric approach to the study of generalized inverse of matrices developed earlier is reviewed and some new results are obtained for applications in the study of linear models.

7. MULTIVARIATE ANALYSIS

7.1 Mixing sequence

The strong law of large numbers is usually proved for a sequence of independent and identically distributed random variables. Recently, some work was done replacing complete independence by pairwise independence. Now the strong law of large numbers is established for a mixing sequence, which is more general than those considered earlier.

7.2 Change point problem

Problems of detecting change points in a process arise in many practical situations. The earlier work done on the change point problem is extended by using rank statistics. Special methods have been developed for detecting changes in the scale and location parameters of directional data.

Information theoretic criteria are used to determine the locations and number of change points, and the strong consistency of these procedures is established. Methods are also devised to detect slope changes.

7.3 Intraclass correlation

Intraclass correlation is defined in situations where measurements are taken on natural clusters of individuals like brothers in a family. A number of problems arise in the study of intraclass correlations. How do we estimate it when observations are available on clusters of different sizes? How do we test the hypothesis that the intraclass correlation is the same in several populations?

The efficiencies of various estimators of the intraclass correlation from sample data have been examined. Tables have been prepared for the percentage points of a number of test criteria for testing the hypothesis of equality of the intraclass correlations.

7.4 Complex multivariate distribution

Several classical tests developed for the real multivariate normal distribution have been extended to complex normal and complex elliptical distributions.

A special study has been made of the various tests concerning the population covariance matrix. Asymptotic distributions have been obtained in each case. The results have wide applicability as they cover important classes of non-normal distributions.

Asymptotic confidence bounds for location parameters, canonical correlations and discriminatory values based on the Fisher discriminant function have been obtained.

7.5 Growth curve model (repeated measurements)

In some practical situations, the structure of Σ, the error covariance in a growth curve model may be known, in which case the estimation of parameters poses new problems.

One case of interest is where Σ has the autoregressive covariance structure. The maximum likelihood estimates of the unknown parameters in this case and their asymptotic distributions are obtained. The likelihood ratio statistic for testing the autoregressive covariance structure is presented.

Another interesting case is where Σ is of the form $X\Gamma X' + \sigma^2 I$. Maximum likelihood estimates of Γ and σ^2 are obtained. Likelihood

ratio tests for hypotheses on other parameters and for the structure of Σ have been derived.

A general linear model with latent variables is considered and the problem of prediction of latent variables and the estimation of all the ancillary unknown parameters are discussed.

APPENDIX

LIST OF TECHNICAL REPORTS AND ABSTRACTS

All the Technical Reports were written with complete or partial support under contract AFSO-88-0030 with the Air Force Office of Scientific Research during the period October 1, 1987-December 31, 1988.

1. Babu, G. Jogesh and Rao, C. Radhakrishna. Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population. Technical Report No. 87-42, Center for Multivariate Analysis, October 1987.

   The joint asymptotic distributions of the marginal quantiles and quantile functions in samples from a p-variate population are derived. Of particular interest is the joint asymptotic distribution of the marginal sample medians, on the basis of which tests of significance for population medians are developed. Methods of estimating unknown nuisance parameters are discussed. The approach is completely nonparametric.

2. Hedayat, A. S., Rao, C. Radhakrishna., and Stufken, J. Designs in survey sampling avoiding contiguous units. Technical Report No. 87-43, Center for Multivariate Analysis, November 1987.

   We review the results on balanced sampling designs excluding contiguous units, as introduced by Hedayat, Rao and Stufken (1987). Some new designs are exhibited, including a design for which $\pi_{ij} = 0$ if $j \equiv i - 2, i - 1, i + 1$ or $i + 2 \pmod N$, and $\pi_{ij} = c$, for a suitable constant c, otherwise. The nonexistence of designs with $N = 3n$, $n \geq 5$, is stated, as well as the uniqueness of the design with $N = 12$, $n = 4$. A discussion on the implementation of the sampling designs obtained through the various constructions is given in the last section.

3. Rao, C. Radhakrishna. A unified approach to estimation in linear models with fixed and mixed effects. Technical Report No. 87-44, Center for Multivariate Analysis, November 1987.

   A unified approach is developed for the estimation of unknown fixed parameters and prediction of random effects in a mixed Gauss-Markoff linear model. It is shown that both the estimators and their mean square errors can be expressed in terms of the elements of a g-inverse of a partitioned matrix which can be set up in terms of the matrices used in expressing the model. No assumptions are made on the ranks of the matrices involved. The method is parallel to the one developed by the author in the case of the fixed effects Gauss-Markoff model using a g-inverse of a partitioned matrix (Rao 1971, 1972, 1973, 1985).

A new concept of generalized normal equations is introduced for the simultaneous estimation of fixed parameters, random effects and random error. All the results are deduced from a general lemma on an optimization problem. This paper is self contained as all the algebraic results used are stated and proved. The unified theory developed in an earlier paper (Rao, 1988) is somewhat simplified.

4. Bai, Z. D., Rao, C. Radhakrishna., and Yin, Y. Q. Least absolute deviations analysis of variance. Technical Report No. 87-45, Center for Multivariate Analysis, November 1987.

Asymptotic methods for testing linear hypotheses based on the L_1-norm regression estimator have been recently discussed by a number of authors. The suggested tests are similar to those based on the least squares theory. Reduction in sums of squares is simply replaced by reduction in sums of absolute deviations. The appropriate distribution theory in such a case has been developed by a number of authors. The object of the present paper is to provide a rigorous proof of the asymptotic distribution of the reduction in sum of absolute deviations, the statistic used in testing a linear hypothesis. The asymptotic distribution is not directly useful as it involves a nuisance parameter. A new method of adjusting for the unknown parameter is suggested.

5. Bai, Z. D., Chen, X. R., Miao, B. Q. and Wu, Y. H. On solvability of an equation arising in the theory of M-estimates. Technical Report No. 87-46, Center for Multivariate Analysis, November 1987.

This article, by otaining the limit of probability that some equation arising in a case of M-estimate possesses at least one solution, establishes the fact that even in the simplest case, when the function ρ is not differentiable at least at one point, it is not legitimate to convert the minimization problem.

6. Chen, X. R., and Wu, Y. H. Strong law for mixing sequence. Technical Report No. 87-47, Center for Multivariate Analysis, December 1987.

In this note we present some theorems on the strong law for the mixing sequence which is not necessarily stationary, and the mixing coefficient involving only a pair of variables in the sequence.

7. Krishnaiah, P. R. and Miao, B. Q. Review about estimation of change point. Technical Report No. 87-48, Center for Multivariate Analysis, June 1987.

This paper gives a detailed survey of the parametric methods and results of statistical inference of change-point models in recent years. The emphasis is on the pure-jump models and segmented linear models, which are dealt with usually by the maximum likelihood and Bayesian methods. Included are various asymptotic results and a short survey of some aspects of nonparametric methods.

8. Bai, Z. D., Subramanyam, K., and Zhao, L. C. On determination of the order of an autoregressive model. Techncial Report No. 87-49, Center for Multivariate Analysis, December 1987.

   To determine the order of an autoregressive model, a new method based on information theoretic criterion is proposed. This method is shown to be strongly consistent and the convergence rate of the probability of wrong determination is established.

9. Bai, Z. D., Subramanyam, K., and Zhao, L. C. Determination of the order of ARIMA process. Technical Report No. 87-50, Center for Multivariate Analysis, December 1987.

   In this paper, using information theoretic criteria: a new method to estimate the order of autoregressive integrated moving average (ARIMA) model is proposed. This procedure yields a strongly consistent estimate of the orders of ARIMA model.

10. Rao, C. Radhakrishna. Weighted and clouded distributions. *Technical Report No. 88-01, Center for Multivariate Analysis, February 1988.*

    The concept of weighted distributions can be traced to the study of effects of methods of ascertainment upon the estimation of frequencies by Fisher in 1934. It was formulated in general terms by the author in a paper presented at the First International Symposium on Classical and Contagious Distributions held in Montreal in 1963. Since then a number of papers have appeared on the subject. This article reviews the previous work and the current developments with some examples.

    Weighted distributions occur in a natural way when adjustments have to be made in the original probability distribution due to deviations from simple random sampling in collecting data, as when the events that occur do not have the same chance of coming into the sample. The examples include: p.p.s. (probability proportional to size) sampling in sample surveys, damage models, visibility bias in quadrat sampling in ecological studies, sampling through effected individuals in genetic studies, waiting time paradox and so on.

11. Miao, B. Q., and Zhao, L. C. Detection of change points using rank methods. *Technical Report No. 88-02, Center for Multivariate Analysis, February 1988.*

    In this paper, the detection and estimation of change points of local parameters are studied by means of localization procedures and rank statistics. These techniques are also applied to detection and estimation of the change points of scale parameters and that of location parameters of directional data.

12. Wu, Y. Discrimination analysis when the variates are grouped and observed in sequential order. *Technical Report No. 88-03, Center for Multivariate Analysis, February 1988.*

Suppose that measurements $x_i = (x_{i1}, \ldots, x_{ij_i})$, $i = 1, \ldots, k$, can be taken on a unit sequentially in that order at the prescribed costs C_i, $i = 1, \ldots, k$. The unit comes from one of the two populations H_1 and H_2, and it is desired to select a population (from these two) from which the unit is supposed to belong to, on the basis of the measurements $x_1, x_2, \ldots$. Given the loss incurred by selecting population H_i when in fact it belongs to H_j, the prior probability p_i of H_i $(i = 1,2)$, and assuming that H_i has the normal distribution $N(\mu_i, V)$, $i = 1,2$ we derive the sequential Bayesian solution of the discrimination problem when μ_1, μ_2 and V are known. When μ_i, V are unknown and must be estimated, we propose a solution which is asymptotic Bayesian with exponential convergence rate.

13. Rao, C. Radhakrishna. Linear transformations, projection operators and generalized inverses-A geometric approach. *Technical Report No. 88-04, Center for Multivariate Analysis, March 1988.*

A generalized inverse of a linear transformation A: $v \to w$, where v and w are finite dimensional vector spaces, is defined using geometric concepts of linear transformations and projection operators. The inverse is uniquely defined in terms of specified subspaces $m \subset v, l \subset w$ and a linear transformation N such that AN = O. Such an inverse which is unique is called the *lm*N-inverse. A Moore-Penrose type inverse is obtained by putting N=O.

Applications to optimization problems when v and w are inner product spaces, such as least squares in a general setting, are discussed. The results given in the paper can be extended without any major modification of proofs to bounded linear operators with closed range on Hilbert spaces.

14. Cacoullos, T., and Papathanasiou, V. Characterizations of distributions by variance points. *Technical Report No. 88-05, Center for Multivariate Analysis, May 1988.*

The distribution of a continuous r.v. X is characterized by the function w appearing in the lower bound $\sigma^2 E^2[w(X)g'(X)]$ for the variance of a function g(X); for a discrete X, $g'(x)$ is replaced by $\Delta g(x) = g(x+1) - g(x)$. The same characterizations are obtained by considering the upper bound $\sigma^2 E\{w(X)[g'(X)]^2\} \geq \mathrm{Var}[g(X)]$. The special case $w(x) = 1$ gives the normal, Borovkov and Utev (1983), and the Poisson, Prakasa Rao and Sreehari (1987). The results extend to independent random variables.

15. Cacoullos, T. On the optimality of the linear discriminant function for spherically isopycnic distributions. *Technical Report No. 88-06, Center for Multivariate Analysis, May 1988.*

The minimum distance (MD), linear discriminant function (LDF), classification rule (CR) is shown to be (a) the minimum Hellinger distance rule and (b) the admissible minimax, symmetric likelihood ratio procedure, for classifying a vector observation X into one of two spherical normal mixtures (SNM) with known location parameters μ_1, μ_2. The normal distribution is characterized by the fact that it maximizes the minimax probability of correct classification in the SNM class with fixed Mahalanobis distance between two alternatives. Some monotone properties and applications are shown for a larger family of spherical distributions (SD). Relations between LDF, CR, Hellinger (affinity) CR and the (admissible) likelihood ratio CR are explored for the k-population case. It is asserted that the LDF, CR are admissible only under a normal SD. A relevant nearest-population problem is also considered.

16. Rao, C. Radhakrishna., and Wu, Y. A strongly consistent procedure for model selection in regression problem. *Technical Report No. 88-07, Center for Multivariate Analysis, May 1988.*

We consider the multiple regression model $y_n = X_n\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}_n$, where y_n and $\underset{\sim}{\varepsilon}_n$ are n-vector random variables, X_n is an n×m matrix and $\underset{\sim}{\beta}$ is an m-vector of unknown regression parameters. Each component of $\underset{\sim}{\beta}$ may be zero or non-zero, which gives rise to 2^m possible models for multiple regression. We provide a decision rule for the choice of a model which is strongly consistent for the true model as $n \to \infty$. The result is proved under certain mild conditions, for instance, without assuming normality of the distribution of the components of $\underset{\sim}{\varepsilon}_n$.

17. Miao, B. Q., Subramanyam, K., and Zhao, L. C. On detection and estimation of change points. *Technical Report No. 88-08, Center for Multivariate Analysis, May 1988.*

Using information theoretic criterion, the problem of change points is considered. In the framework of model selection, procedures are developed to estimate the locations and the number of change points. These procedures are shown to be strongly consistent in estimating the number and location of change points in the mean vector when the covariances are different.

18. Bai, Z. D., Chen, X. R., Miao, B. Q., and Rao, C. Radhakrishna. Asymptotic theory of least distances estimate in multivariate linear models. *Technical Report No. 88-09, Center for Multivariate Analysis, May 1988.*

We consider the multivariate linear model

$$Y_i = X_i'\beta_0 + \epsilon_i, \qquad i = 1, \ldots, n$$

where Y_i is a p-vector random variable, X_i is a q×p matrix, β_0 is

an unknown q-vector parameter and $\{\epsilon_i\}$ is a sequence of iid p-vector random variable with median vector zero. The estimate $\hat{\beta}_n$ of β_0 such that

$$\min_{\beta} \sum_{i=1}^{n} \|Y_i - X_i'\beta_0\| = \sum_{i=1}^{n} \|Y_i - X_i'\hat{\beta}_n\|$$

is called the least distances (LD) estimator. It may be recalled that the least squares (LS) estimator is obtained by minimizing the sum of norm squares.

In this paper, it is shown that the LD estimator is unique, consistent and has an asymptotic q-variate normal distribution with mean β_0 and covariance matrix V which depends on the distribution of the error vectors $\{\epsilon_i\}$. A consistent estimator of V is proposed which together with $\hat{\beta}_n$ provide an asymptotic inference on β_0. In particular, tests of linear hypotheses on β_0 analogous to those of analysis of variance in the Gauss-Markoff linear model are developed. Explicit expressions are obtained in some cases for the asymptotic relative efficiency of the LD compared to the LS estimator.

19. Rao, B. Raja and Talwalker, Sheela. 'Setting the clock back to zero' property of a life distribution. *Technical Report No. 88-10, Center for Multivariate Analysis, May 1988.*

In the present paper, we have developed a general class of life distributions, following Krane's (1963) assumption that a polynomial of degree m of the life length X of an item, that is, the random variable $y(X) = \beta_1 X + \beta_2 X^2 + \ldots + \beta_m X^m$, follows an exponential distribution with mean unity. Such a class of life distributions, has a remarkable property, called, 'Setting the clock back to zero' property. This property ensures that the form of the life distribution remains unchanged, except for some parameter values, when the population of individuals who have survived a given period of time x_0 is considered, together with a transformation $X_1 = x - x_0$, so that $X_1 \geq 0$. The advantage of having such a property is in the area of many epidemiological, biomedical and engineering experiments, in which truncated data are very common. The problems of estimation, confidence intervals and testing hpotheses are greatly simplified.

20. Miao, B. Q. and Subramanyam, K. On some methods of estimation of slope change points. *Technical Report No. 88-11, Center for Multivariate Analysis, May 1988.*

Change points can be classified into two types: jump change and slope change. In this paper, a procedure to detect and estimate the number and locations of slope change points is presented. This procedure gives strongly consistent estimates. This method can be extended to

multivariate case easily.

21. Subramanyam, K. and Rao, M. B. On the structure of $2\times\infty$ bivariate distributions which are totally positive of order two. *Technical Report No. 88-12, Center for Multivariate Analysis, June 1988.*

Let X and Y be two real random variables such that X takes only two values 1 and 2. The notion of total positivity of order two for the joint probability density function of X and Y is discussed in this paper from the viewpoint convex analysis.

22. Khatri, C. G. and Cacoullos, T. Characterization of distributions within the elliptical class by a Gamma distributed quadratic form. *Technical Report No. 88-13, Center for Multivariate Analysis, June 1988.*

Let x be spherically distributed with characteristic function $\phi(t't)$ for all $t \in R^n$, and let $x'Ax$ be a quadratic form where A is a symmetric matrix of rank m ($\leq n$). Assume that the density of x exists and is infinitely differentiable. Then $x'Ax \sim G(a,\theta)$, $a > 0$, $\theta > 0$ if and only if $A^2 = \lambda A$ for some $\lambda(> 0)$ and

$$\phi(t't) = {}_1F_1(a;\tfrac{1}{2}m;\ -t't/4\theta\lambda,\ t \in R^n.$$

If $a = \frac{1}{2}m$, then we get the normality of x while if $m = n$, the density of x is given by

$$\{(\lambda\theta)^a\Gamma(\tfrac{1}{2}n)/\Gamma(a)\pi^{n/2}\}(x'x)^{a-n/2}\exp(-\theta\lambda(x'x)),\ x \in R^n.$$

Here, $G(a,\theta)$ denotes the Gamma-distribution whose density function is given by

$$\{\theta^a/\Gamma(a)\}Z^{a-1}\exp(-\theta Z) \quad \text{for all} \quad Z > 0.$$

This corrects the characterization of normality as given by Khatri and Mukerjee (1987). This result is extended for matrix spherical, matrix elliptical, complex elliptical and matrix complex elliptical variates.

23. Gupta, Shanti S, and Liang, TaChen. On a sequential subset selection procedure. *Technical Report No. 88-14, Center for Multivariate Analysis, June 1988.*

This paper deals with the problem of selecting the best population through the sequential subset selection approach. Based on the modified likelihood ratio of the probability density function of some invariant sufficient statistics, a sequential subset selection procedure is proposed. When the procedure terminates, one can assert with a guaranteed probability P*, that the best population is included in the selected subset and that each selected population is within some fixed distance from the best population.

24. Khatri, C. G. Study of redundancy of vector variables in canonical correlations. *Technical Report No. 88-15, Center for Multivariate*

Analysis, June 1988.

Fujikoshi (1982) obtained the necessary and sufficient conditions for the increased number of variables in the two sets of vectors not affecting the original nonzero canonical correlations and used these to obtain the likelihood ratio test procedure. He assumed a nonsingular covariance matrix due to random variables. Here, we study the same problem when the covariance matrix is singular and establish some further results. In this study, we note that the unit canonical correlations have to be separated in some of the situations. These

results are valid for complex random vector variables and in some situations, the test for redundancy is given for complex random variables.

25. Fujikoshi, Yasunori. Error bounds for asymptotic expansions of the multivariate t- and F-variables with common denominator. *Technical Report No. 88-16, Center for Multivariate Analysis, June 1988.*

Let $X = (X_1,\ldots,X_p)$ be a scale mixture of a p-dimensional random vector $Z = (Z_1,\ldots,Z_p)$ with scale factor $\sigma > 0$, i.e., $X = \sigma Z$, where Z and σ are independent. We are concerned with asymptotic expansions of the distribution function of $\mathrm{Max}(X_1,\ldots,X_p)$ in the two cases:

(i) $Z_1,\ldots,Z_p$ i.i.d. $\sim N(0,1)$, $\sigma = (\chi^2_n/n)^{1/2}$, (ii) $Z_1,\ldots,Z_p$ i.i.d. $\sim$ $G(\lambda)$, $\sigma = \chi^2/n$. We give a unified derivation of the asymptotic expansions as well as their error bounds.

26. Khatri, C. G and Bhavsar, C. D. Some asymptotic inferential problems connected with complex elliptical distribution. *Technical Report No. 88-17, Center for Multivariate Analysis, June 1988.*

The paper extends the results of Khatri (1988) to complex elliptical variates. Asymptotic confidence bounds on location parameters for the linear growth curve for the complex variates, the asymptotic distribution of the canonical correlations for the two sets of complex variates and the asymptotic confidence bounds for the discriminatory values (see Khatri et all, 1986) for the linear Fisher's discriminator for the future complex observation z are developed in this paper on the lines given by Khatri (1988).

27. Rao, B. Raja and Talwalker, Sheela. Bounds on the life expectancy for the Rayleigh and the Weibull distributions. *Technical Report No. 88-18, Center for Multivariate Analysis, July 1988.*

The present paper gives bounds on the life expectancy or the mean residual life of an individual, whose life length is a random variable X following a Rayleigh distribution, or more generally a Weibull distribution. Simple transformations of the variables give inequalities on the Mills' ratio and the incomplete gamma functions.

Some numerical computations are also reported to compare the lower and upper bounds with the exact value of the life expectancy function for several values of the parameter.

28. Fujikoshi, Y., Kanda, T. and Tanimura, N. The growth curve models with an autoregressive covariance structure. *Technical Report No. 88-19, Center for Multivariate Analysis, July 1988.*

The growth curve model with an autoregressive covariance structure is considered. An iterative algorithm for finding the MLE's of the parameters in the model is presented, based on the modified likelihood equations. Asymptotic distributions of the MLE's are obtained when the sample size is large. The likelihood ratio statistic for testing the autoregressive covariance structure is presented.

29. Khatri, C. G. Some properties of BLUE in a linear model and canonical correlations associated with linear transformations. *Technical Report No. 88-20, Center for Multivariate Analysis, July 1988.*

Let $(x, X\beta, V)$ be a linear model and let $A' = (A_1', A_2')$ be a $p \times p$ nonsingular matrix such that $A_2X = 0$, Rank A_2 = p - Rank X. We represent the BLUE and its covariance matrix in alternative forms under the condition that the number of unit canonical correlations between $y_1 (= A_1x)$ and $y_2 (= A_2x)$ is zero. For the second problem, let $x' = (x_1', x_2')$ and let a g-inverse V^- of V be written as $(V^-)' = (A_1', A_2')$. We investigate the relations (if any) between the nonzero canonical correlations $\{1 \geq \rho_1 \geq \ldots \geq \rho_t > 0\}$ due to $y_1 (= A_1x)$ and $y_2 (= A_2x)$, and the nonzero canonical correlations $\{1 \geq \lambda_1 \geq \ldots \geq \lambda_{v+\epsilon} > 0\}$ due to x_1 and x_2. We answer some of the questions raised by Latour, et al (1987) in the case of the Moore-Penrose inverse $V^+ = (A_1', A_2')$ of V.

30. Rao, M.B. and Velu, R. On inferences about interclass correlations from familial data. *Technical Report No. 88-21, Center for Multivariate Analysis, July 1988.*

The main objectives of this paper are:

1. To compare the bias and mean square error or Srivastava's Ensemble estimators;
2. To derive the exact distribution of Sib-Mean estimator under the hypothesis that the population interclass correlation is zero;
3. To derive the exact distributions of Srivastava's and Ensemble estimators under the hypothesis that the population interclass correlation is zero;
4. To present a Monte Carlo study of Srivastava's estimator in testing of hypotheses.

31. Bhavsar, C. D., and Khatri, C. G. Asymptotic distributions of test statistics for covariance matrices concerning complex elliptical distributions. *Technical Report No. 88-22, Center for Multivariate Analysis, July 1988.*

Let x be a complex random vector and let it have a complex elliptical distribution $E_p(\mu,\Sigma;\Psi)$. The various tests of hypotheses concerning Σ similar to the problems on the real case developed by Khatri and Bhavsar (1988b) are considered and their asymptotic distributions of the likelihood ratio tests obtained under normality assumption are established for the complex elliptical class of distributions. These asymptotic distributions are either non-central chi-squares or that of a linear function of non-central chi-square variates.

32. Rao, M. Bhaskara. On the matching problem. *Technical Report No. 88-23, Center for Multivariate Analysis, July 1988.*

In a random distribution of n balls numbered from 1 to n into n cells numbered from 1 to n so that each cell receives exactly one ball, a match is said to occur if a ball bearing a certain number goes into the cell bearing the same number. The distribution of the number of matches is well known. In this article, an elementary argument is presented to derive this distribution based on a certain recurrence property. This argument helps to derive all the moments of the distribution of the number of matches.

33. Baksalary, Jerzy K., Liski, Erkki P., and Trenkler, Gotz. Mean square error matrix improvements and admissibility of linear estimators. *Technical Report No. 88-24, Center for Multivariate Analysis, July 1988.*

In the first part of this paper, the set $L(Cy+c)$ comprising all linear estimators of β which are as good as a given unbiased estimator $Cy + c$ with respect to the mean square error matrix criterion in at least one point of the parameter space is investigated under the unrestricted linear regression model $M = \{y,X\beta,\sigma^2 I_n\}$ and the restricted model $M_0 = \{y,X\beta|R_0\beta = r_0,\sigma^2 I_n\}$. In the second part, new characterizations of the sets A and A_0 of all linear estimators that are admissible for β under M and M_0 with respect to the mean square error criterion are derived referring to the sets $L(\hat{\beta})$ and $L(\hat{\beta}_0)$, where $\hat{\beta}$ and $\hat{\beta}_0$ are the minimum dispersion linear unbiased estimators of β in these two models. The convexity of the sets $L(Cy+c)$, A and A_0 is also pointed out.

34. Rao, C. Radhakrishna and Shanbhag, D. N. Recent Advances on the integrated Cauchy functional equation and related results in applied probability. *Technical Report No. 88-25, Center for Multivariate Analysis, July 1988.*

The integrated cauchy functional equation appears in several

characterization problems in applied probability. This is evident from Lau and Rao (1982), Rao and Shanbhag ((1986), (1987)), and Davies and Shanbhag (1987) among others. Various general results on the equation have been given by Choquet and Deny (1960), Deny (1961), Davies and Shanbhag (1987) and Rao and Shanbhag (1987). The present paper aims at reviewing these results with improvements wherever possible. Some further applications of these results in applied probability are also discussed.

35. Alzaid, Abdulhamid A., Rao, C. Radhakrishna, and Shanbhag, D. N. Elliptical symmetry and exchangeability with characterizations. *Technical Report No. 88-26, Center for Multivariate Analysis, July 1988.*

In this paper we establish certain general characterization results on elliptically symmetric distributions and exchangeable random variables. These results yield in particular the results given earlier by Maxwell (1960), Bartlett (1934), Kingman (1972), Ali (1980), Smith (1981), Arnold and Lynch (1982) and several others as straight forward corollaries.

36. Kagan, Abraham. The Lukacs-King method applied to problems involving linear forms of independent random variables. *Technical Report No. 88-27, Center for Multivariate Analysis, July 1988.*

Exposed in the paper are some recent results, including a few new results on linear forms of independent random variables obtained by a method first used in Lukacs and King (1954). Though the result explicitly formulated in this paper is weaker than the well known Darmois-Skitovitch theorem proved by Darmois and Skitovitch independently of each other and of Lukacs and King and published at about the same time, the method of the above paper actually proves a stronger result than formulated and can be applied to other characterization problems in terms of linear forms of independent random variables.

37. Sambamoorthi, N. Information theoretic criterion approach to dimensionality reduction in multinomial logistic regression models. Part I: Theory. *Technical Report No. 88-28, Center for Multivariate Analysis, July 1988.*

We discuss the issue of dimensionality reduction in multinomial logistic regression models as problems arising in variable selection, collapsibility of responses and linear restrictions in the parameter matrix. A method using information theoretic criterion suggested by Bai, Krishnaiah and Zhao (1987), which is a variant of Akaike Information Criterion (AIC), is used to estimate the rank of the parameter matrix. The same procedure is used for the selection of variables and the collapsibility of response categories. This technique yields strongly consistent estimates, whereas AIC fails to provide consistent estimates.

38. Sambamoorthi, N. Information theoretic criterion approach to

dimensionality reduction in multinomial logistic regression models. Part II: Simulations. *Technical Report No. 88-29, Center for Multivariate Analysis, July 1988.*

In Part I, we proposed an information theoretic criterion for (1) identification of the rank of the parameter matrix, (2) selection of variables, and (3) collapsibility of response categories in multinomial logistic regression models. The proposed procedure gives strongly consistent estimates. It is important to see the efficacy of such procedures for moderate sample sizes. In this paper, we report the simulation results of variable selection problem. The results show that if we choose the criterion function suitably, then the probability of misidentification could be significantly lower than the Akaike Information Criterion even for small sample sizes. Thus, if minimization of probability of misidentification is a useful goal, then the proposed procedure is preferable. The problem of exactly identifying the criterion function which has the lowest probability of misidentification is still open.

39. Kagan, Abram and Rao, C. Radhakrishna. Constancy of regression of a polynomial of sample average on residuals characterizes normal distribution. *Technical Report No. 88-30, Center for Multivariate Analysis, July 1988.*

Let $X_1,\ldots,X_n$ be iid observations from a distribution function F and $P(\bar{X}) = a_k\bar{X}^k + \ldots + a_0, a_k \neq 0$ be an arbitrary polynomial of degree $k \geq 2$ in $\bar{X}$, the sample average. It is proved that if $n \geq 2k$ and $\alpha_{k+1} = E|X_1|^{k+1} < \infty$ then

$$E(P(\bar{X})|X_1 - \bar{X},\ldots,X_n - \bar{X}) = c(\text{constant})$$

if and only if F is Gaussian. If $P(\bar{X})$ is nonnegative with probability 1, then the condition $\alpha_{k+1} < \infty$ can be weakened to the minimal necessary condition $\alpha_k < \infty$. The case of $k = 1$ was investigated in Kagan, Linnik and Rao (1965) under the conditions $n \geq 3$ and $E|X_1| < \infty$.

40. Baksalary, Jerzy K. and Mathew Thomas. Rank invariance criterion and its application to the unified theory of least squares. *Technical Report No. 88-31, Center for Multivariate Analysis, July 1988.*

Necessary and sufficient conditions are established for the product AB^-C to have its rank invariant with respect to the choice of a generalized inverse B^-. In particular cases, these conditions coincide with the results of Mitra (1972). They are discussed also in the statistical

context of the unified theory of least squares introduced by Rao (1971).

41. Baksalary, Jerzy K. and Markiewicz, Augustyn. Admissible linear estimators of an arbitrary vector of parametric functions in the general Gauss-Markov model. *Technical Report No. 88-32, Center for Multivariate Analysis, July 1988.*

This paper derives a complete charaterization of estimators that are admissible for any, not nessarily identifiable, vector of parametric functions among the set of linear estimators under the general Gauss-Markov model $M = \{Y, X\beta, \sigma^2 V\}$ with both the model matrix X and the dispersion matrix V possibly deficient in rank. This characterization is then applied to examine admissibility of various estimators of β proposed in the literature.

42. Baksalary, Jerzy K., Puntanen, Simo, and Styan, George P. H. A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss-Markov model. *Technical Report No. 88-33, Center for Multivariate Analysis, July 1988.*

Solutions are derived to three different versions of the problem: when the dispersion matrix of the best linear unbiased estimator of the expectation vector in the general Gauss-Markov model can be expressed in a form characteristic for the usual least-squares theory. A common denominator for all those versions is a certain property of the canonical correlations between the vector of the ordinary least-squares fitted values and the vector of the residuals. Among preliminaries, a brief survey of various representations of the dispersion matrix of the best linear unbiased estimator is given, as well as some auxiliary algebraic results that seem to be of interest also independently of the statistical context.

43. Baksalary, Jerzy K. and Puri, P. D. Pairwise balanced, variance-balanced, and resistant incomplete block designs revisited. *Technical Report No. 88-34, Center for Multivariate Analysis, July 1988.*

A general solution is derived to the problem of characterizing block designs that are simultaneously pairwise- and variance-balanced. Applications of the characterizations obtained to some problems concerned with the local resistance of BIB designs are presented.

44. Babu, Gutti Jogesh and Rao, C. Radhakrishna Estimation of the reciprocal of the density quantile function at a point. *Technical Report No. 88-35, Center for Multivariate Analysis, July 1988.*

Consistent estimators for the reciprocal of the density at a quantile point are considered. Optimal rates of covergence of these estimators, depending on the smoothness properties of the density, are obtained. Two different, but natural, estimators of the reciprocal of the density at a quantile point, based on several samples from a location parameter family with unknown and possibly different location paremeters are proposed. A linear combination of estimates based on individual

samples is shown to be better than the estimate based on pooled samples in the mean squared error sense.

45. Bai, Z.D., Miao, B.Q. and Rao, C. Radhakrishna. Estimation of direction of arrival of signals asymptotic results. *Technical Report No. 88-36, Center for Multivariate Analysis, August 1988.*

A new method is proposed for the estimation of the unknown directions of arrival of signals from various sources. It is suggested that the number of signals be estimated first by using model selection criteria such as those introduced by Bai, Krishnaiah and Zhao, and the estimates of directions of arrival for a given number sources be obtained next. The new method uses the eigen structure property of the covariance matrix, specially of the noise eigen space, a more direct way than in the other proposed algorithms for estimation.

The strong consistency of the estimation has been established and the asymptotic distribution of the estimators has been derived.

46. Srivastava, M.S. Multiple regression method in opthalmology and familial data. *Technical Report No. 88-37, Center for Multivariate Analysis, August 1988.*

Rosner (1984) consider multiple regression method to analyze ophthalmology data and provided an iterative solution using Newton-Raphson method. In this paper an explicit solution is given without the assumption of normality. Also, an exact test for the significance of the intraclass correlation is presented.

47. Srivastava, M.S. and Yau, Wai Kwok. Tail probability approximations of a general statistics. *Technical Report No. 88-38, Center for Multivariate Analysis, August 1988.*

Two explicit approximation formulae for the tail probability of a general statistic are derived. The observations on which the general statistic is based need not be identically distributed or even independent. The first one is based on the Edgeworth expansion of the exponentially shifted density recentered at the value of the statistic as in Robinson (1982) and Daniels (1987). The second one uses Bleistein's (1966) idea in dealing with a saddlepoint near a simple pole at the origin as in Lugannani and Rice (1980). Illustrative examples include, the tail probability of the sum of independent noncentral chi-square random variables, Durbin-Watson statistics, and linear combination of noncentral chi-square random variables.

48. Dahiya, Ram C. and Hossain, Syed A. Estimating the parameters of a non-homogeneous poisson process model for software reliability. *Technical Report No. 88-39, Center for Multivariate Analysis, August 1988.*

A stochastic model for the software failure phenomenon based on a nonhomogeneous Poisson process (NHPP) was suggested by Goel and Okumoto (1979). The model has been widely used but very little work has been

done on the problem of estimating the parameters. We present a necessary and sufficient condition for the likelihood estimates to be finite, positive and unique. The probability distribution of faults remaining after debugging and the problem of estimating the expected number of remaining faults are investigated here. The results obtained here are applied to two real life examples pertaining to software failure data.

49. Khatri, C. G. and Bhavsar, C. D. Asymptotic distributions of test statistics for covariance matrices concerning elliptical distributions. *Technical Report No. 88-40, Center for Multivariate Analysis, August 1988.*

This article presents explicitly the results on the asymptotic distributions of the likelihood ratio test statistic $-2 \log \lambda$ ($= n\hat{F}$) when the sampling is from the nonnormal populations possessing the first four moments similar to those of an elliptically contoured distribution. The statisticso $\hat{F}$ are obtained on the various structures of Σ for one or more populations. All the situations, the asymptotic distributions of $n\hat{F}$ are either noncentral Chi-squares or those of a linear function of two noncentral Chi-square variates, when the alternatives are close to null hypotheses. For other alternatives, we get asymptotic normality of $\sqrt{n}(F-F_0)/\sigma_0$ where $\sqrt{n}\ E(\hat{F}) = \sqrt{n}\ F_0 + 0(1)$ and $V(\hat{F}) = \sigma_0^2/n + 0(n^{-2})$.

50. Khatri, C.G., Pukkila, T.M. and Rao, C. Radhakrishna. Tables for testing intraclass correlation coefficients. *Technical Report No. 88-41, Center for Multivariate Analysis, August 1988.*

Tables for one-sided, two-sided unbiased and likelihood ratio tests for testing equality of intraclass correlations for two multivariate normal populations are prepared for $p = 2,3,4,5$ and $n_1, n_2 =$ $4,5,6,7,8,9,10,12,15,20,25,30,40,60,120,999$. By simulations, it is shown that the likelihood ratio test for testing the equality of two intraclass correlations for unequal ρ_1 and ρ_2 variates normal populations appears to depend on the nuisance parameter ρ, the common intraclass correlation under H_0 when the sample sizes are small. The one degree of freedom chisquare approximation to the likelihood ratio test statistic is sufficiently accurate for all values of ρ when sample sizes are over 20, and could be used in practice even in small samples although it overestimates significance.

51. Rao, C. Radhakrishna. Methodology based on the L_1-norm in statistical inference. *Technical Report No. 88-42, Center for Multivariate Analysis. September 1988.*

The paper reviews some recent contributrions to statistical methodology based on the L_1-norm as a robust alternative to that based on the least squares. Tests are developed using the medians instead of the means

and least absolute deviations instead of least squares. Analogues of of Hotelling's T^2 and tests based on the roots of a determinatal equation are derived using medians.

Asymptotic inference procedures on regression parameters in the univariate linear model are reviewed and some suggestions are made for the elimination of nuisance parameters which occur in the asymptotic distributions. The results are extended to the multivariate linear model.

Recent work on the asymptotic theory of inference on the parameters of a generalized multivariate linear model based on the method of least distances is discussed. New tests are developed using least distances estimators.

52. Baksalary, Jerzy K., Rao, C. Radhakrishna and Markiewicz, Augustyn. A study of the influence of the "natural restrictions" on estimation problems in the singular Gauss-Marov model. *Technical Report No. 88-43, Center for Multivariate Analysis, October 1988.*

It is known that if the Gauss-Markov model $M = \{Y, X\beta, \sigma^2 V\}$ has the column space of the model matrix X not contained in the column space of the dispersion matrix V, then the vector of parameters β has to satisfy certain linear equations. However, these equations become restrictions on β in the usual sense only when the random vector Y occurring in them is replaced by an observed outcome y. In this paper, explicit solutions to several statistical problems are derived in two situations: when β is unconstrained and when β is constrained by two "natural restrictions" mentioned above. The problems considered are: linear unbiased estimation and best linear untiased estimation of an identifiable vector of parametric functions, comparison of estimators of any vector of parametric functions with respect to the matrix risk, and admissibility among the class of all linear estimators with respect to the matrix risk and with respect to the mean square error. The solutions corresponding to the unconstrained and constrained cases are compared to show in what sense β may be considered to be free to vary without loss of generality.

53. Rao, B. Raja, Talwalker, Shella and Kundu, Debsis. Confidence intervals for the relative risk ratio parameter from survival data under a random epidemiologic studies. *Technical Report No. 88-44, Center for Multivariate Analysis, October 1988.*

The present paper reports the results of a Monte Carlo simulation study to examine the performance of several approximate confidence intervals for the Relative Risk Ratio (RRR) parameter in an epidemiologic study, involving two groups of individuals. The first group consists of n_1 individuals, called the experimental group, who are exposed to some carcinogen, say radiation, whose effect on the incidence of some form of cancer, say skin cancer, is being investigated. The second group consists of n_2 individuals (called the control group, who are exposed

to the carcinogen. Two cases are considered in which the life times (or time to cancer) in the two groups follow (i) the expondential and (ii) the Weibull distributions. The case when the life times follow a Rayleigh distribution follows as a particular case. A general random censorship model is considered in which the life times of the individuals are censored on the right by random censoring times following (i) the exponential and (ii) the Weibull distributions. The Relative Risk Ratio parameter in the study is defined as the ratio of the hazard rates in the two distributions of the times to cancer. Approximate confidence intervals are constructed for the RRR parameter using its maximum likelihood estimator (m.l.e.) and several other methods, including a method due to Fieller. Sprott's (1973) and Cox's (1953) suggestions, as well as the Box-Cox (1964) transformation, are also utilized to construct approximate confidence intervals. The performance of these confidence intervals in small samples is investigated by means of some Monte Carlo simulations based on 500 random samples. Our simulation study indicates that many of these confidence intervals perform quite well in samples of size 10 and 15, in terms of the coverage probability and expected length of the interval.

54. Babu, Gutti Jogesh. Strong representations for LAD estimators in linear models. *Technical Report No. 88-45, Center for Multivariate Analysis, October 1988.*

Consider the standard linear model $y_i = z_i\beta + e_i$, i=1, 2, ..., n, where z_i denotes the ith row of an n×p design matrix, $\beta\epsilon\mathbb{R}^p$ is an unknown parameter to be estimated and e_i are independent random variables with a common distribution function F. The least absolute deviation (LAD) estimate $\hat{\beta}$ of β is defined as any solution of the minimization problem

$$\sum_{i=1}^{n} |y_i - z_i\hat{\beta}| = \inf\{ \sum_{i=1}^{n} |y_i - z_i\beta| : \beta\epsilon\mathbb{R}^p\}.$$

In this paper Bahadur type representations are obtained for $\hat{\beta}$ under very mild conditions on F near zero and on z_i, i=1, ..., n. These results are extended to the case, when $\{e_n\}$ is a mixing sequence. In particular the results are applicable when the residuals e_i form a simple autoregressive process.

55. Khatri, C. G. Multivariate generalization of t'-statistic based on the mean square successive difference. *Technical Report No. 88-46. Center for Multivariate Analysis, October 1988.*

The usual t-statistic is not useful if the successive observations have some kind of linear trend. This generally arises in the drug testing experiment and it is clearly pointed by Shah (1988). He suggests to use

t'-statistic which is defined by $t' = \sqrt{k}\,\bar{y}/\delta_1$ where $\bar{y} = \sum_{i=1}^{n} y_i/n$, $\delta_1^2 = \sum_{i=1}^{k-1} (y_{i+1}-y_i)^2/2(k-1)$ and $y_1,\ldots,y_k$ are independent observations from $N(\mu,\sigma^2)$. We generalize this statistic to multivariate situation and define T'-statistic as $T' = k\bar{\underset{\sim}{y}}'\Delta_1^{-1}\bar{\underset{\sim}{y}}$ where $\Delta_1 = \sum_{i=1}^{k-1} (\underset{\sim}{y}_{i+1}-\underset{\sim}{y}_i)(\underset{\sim}{y}_{i+1}-\underset{\sim}{y}_i)'/2(k-1)$. The exact null distribution of T' and an approximate null distribution of T' are obtained. For $p = 1$, this approximate values are compared with the exact values of t' at 5% level. The approximation is found to be appropriate for all practical purposes.

56. Khatri, C. G., Khattree, Ravindra and Gupta, Rameshwar D. On a class of orthogonal invariant and residual independent matrix distributions. *Technical Report No. 88-47, Center for Multivariate Analysis, October* 1988.

Let X and Y be independent positive definite random matrices and let their distributions belong to the class C of the Orthogonal Invariant and Residual Independent Matrix (Oriarim) distributions. Let T be any square root of Y in the sense $Y = TT'$ for the real random matrix Y (or $Y = TT^*$ for the complex random matrix Y with T^* being a conjugate transpose of T). Then, the distribution of TXT' (or TXT^*) is Oriarim and belong to C. Some special distribtions useful to signal detection are given to point out the importance of this class C.

57. Khatri, C. G. and Rao, C. Radhakrishna. Multivariate linear model with latent variables: problems of estimation. *Technical Report No. 88-48, Center for Multivariate Analysis, November 1988.*

Consider a linear model

$$Y_i = X_i\beta_i + Z_i\gamma + \epsilon_i,\ \beta_i = Cu_i + \eta_i,\ i = 1,\ldots,n$$

where β_i are latent vector variables, and ϵ_i, η_j are error vector variables such that

$$E(\epsilon_i) = 0,\ D(\epsilon_i) = \sigma^2 I,\ E(\eta_j) = 0,\ D(\eta_j) = \Gamma.$$

Such a model arises in problems of selection based on an inherent quality of an individual, which is not directly observable. The problems discussed in this paper are the estimation of the unknown parameters γ, C, σ^2 and Γ, prediction of the latent variables

β_i, i = 1,...,n, for the observed individuals and the prediction of β for a future individual based on the measurement u only.

58. Kagan, Abram. Constancy of regression of a polynomial of sample average of positive random variables on their ratios characterizes gamma distribution. *Technical Report No. 88-49, Center for Multivariate Analysis, November 1988.*

Let $X_1, X_2, \ldots, X_n$ be i.i.d. positive random variables with a distribution function $F(x)$ and $P(\bar{X}) = A_k \bar{X}^k + \ldots + A_0$, $A_k \neq 0$ be a polynomial of degree $k \geq 2$ in $\bar{X}$, the sample average. It is proved that if $n \geq 2k$ and

$$\int_0^\infty x^{k+\epsilon} dF < \infty, \quad \int_0^\infty x^{-\epsilon} dF < \infty$$

for an $\epsilon > 0$, then

$$E\{P(\bar{X}) \mid X_1/\bar{X}, \ldots, X_n/\bar{X}\} = \text{constant}$$

if and only if F is gamma. The case of $k = 1$ was investigated by Khatri and Rao (1968) under the minimal necessary conditions $n \geq 3$ and $E(X_1) < \infty$.

If $F(x)$ contains a scale parameter $\sigma > 0$, $F(x) = F(x/\sigma)$, and $P(\bar{X})$ is used as an unbiased estimator of the parameter polynomial $\pi(\sigma) = E_\sigma P(\bar{X}) = A_k \sigma^k + \ldots + A_0$, then under the conditions

$$\int_0^\infty x^{2k} dF < \infty, \quad \int_0^\infty x^{-\epsilon} dF < \infty$$

for an $\epsilon > 0$, $P(\bar{X})$ is the best unbiased estimator of $\pi(\sigma)$ with respect to quadratic loss if and only if F is gamma.

59. Fujikoshi, Y. and Khatri, C. G. A study of redundancy of some variables in covariate discriminant analysis. *Technical Report No. 88-50, Center for Multivariate Analysis, December 1988.*

Test for redundancy of some variables in discriminant analysis were developed by Rao (1946, 1948), which were further studied by McKay (1977) and Fujikoshi (1982). These tests are now extended to the most general situation which includes redundancy in covariate as well as main variables in discrimination between two or more groups. The likelihood ratio test is derived under multivariate linear and growth curve models. As the asymptotic distribution of the likelihood ratio test is complicated, some alternative methods of testing are suggested.